Frontiers in Economics and Management Volume 5 Issue 4, 2024
ISSN: 2692-7608 DOI: 10.6981/FEM.202404_5(4).0007

A Clustering Location Method Considering the Potential Value of Medical Customers

Yuyuan Yi[1], Weibin Deng[1,*], and Yiming Zhang[2]

[1] Chongqing Smart Posts Engineering Technology Research Center, Chongqing University of Posts and Telecommunications, Chongqing, China

[2] 78111 troops, People's Liberation Army of China, Chengdu, China

*Corresponding Author Email: dengwb@cqupt.edu.cn

Abstract

Aiming at the location problem of distribution center under the condition of uncertain customer value, a K-means clustering location method based on entropy weight method was proposed for pharmaceutical logistics enterprises. Firstly, a set of pharmaceutical customer potential value evaluation index system was constructed from five aspects: customer current value, customer development potential, customer operating cost, regional customer demand and industry competition. Then, the entropy weight method was used to calculate customer potential value, and a CKM clustering algorithm was proposed to constrain potential customer value capacity. ARI, AMI, SC and DBI were used to test the effectiveness of the algorithm. Finally, cluster analysis was carried out on all existing and potential customer data of P pharmaceutical logistics enterprises except the distribution of central warehouse, and cluster partitions and distribution center location points were obtained, which provided powerful decision support for logistics optimization of enterprises.

Keywords

Pharmaceutical Logistics; Distribution Center Location; Customer Value; K-means Clustering.

1. Introduction

At present, China has become the world's second largest pharmaceutical consumer market after the United States. From 2012 to 2023, the sales scale of China's top 100 chain pharmacies increased from 43.9 billion yuan to 297.6 billion yuan, and at the same time, the number of pharmacies climbed from 424,000 to 655,000, achieving rapid growth. In 2023, the drug sales scale of China's pharmaceutical retail market has reached 501.5 billion yuan, an increase of 3.3%. This has also driven the development of the pharmaceutical logistics industry. According to China's regulations on drug sales, drugs must be centrally purchased from manufacturers through pharmaceutical wholesale enterprises and then sold to retailers. Under this special supply chain model, a logistics model dominated by third-party pharmaceutical logistics enterprises is derived, and professional third-party logistics is solely responsible for the service demand of the entire pharmaceutical supply chain[1].

Pharmaceutical logistics enterprise refers to the logistics company focusing on the pharmaceutical industry, its core business is to integrate pharmaceutical sales, storage and transportation and other links, to achieve the efficient, safe and reliable circulation of pharmaceutical products for downstream pharmaceutical retail customers, to provide a strong logistics guarantee for the entire pharmaceutical industry. Among them, downstream pharmaceutical retail customers (referred to as pharmaceutical customers) include pharmacies, large public hospitals, community hospitals, rural clinics and personal clinics. For

ISSN: 2692-7608 DOI: 10.6981/FEM.202404_5(4).0007

pharmaceutical logistics enterprises, scientific and reasonable selection of distribution center location and optimization of vehicle distribution route is the key to improve the efficiency of the entire drug distribution network. Since the construction of distribution centers is often characterized by large costs and long cycles, it is necessary to consider not only the needs of existing customers, but also the conversion of potential customers when selecting distribution centers.

With the continuous expansion of the number of customers and business coverage, the original central warehouse is difficult to meet the distribution needs of customers, and many pharmaceutical logistics enterprises have established distribution centers to provide drug distribution services for customers. At the same time, faced with the task of online procurement transformation for pharmaceutical customers, third-party pharmaceutical logistics enterprises are scrambling for the market. In addition to existing customers, other pharmaceutical customers within the scope of business will become potential customers of enterprises [2]. It is an important task for sales personnel of logistics enterprises to explore potential medical customers and convert them into new customers [3]. Therefore, when establishing a distribution center, it is difficult to maximize the utilization of resources only considering the needs of existing customers. How to scientifically evaluate the value of potential customers and consider them into the entire logistics demand network through clustering has become an important topic for many enterprises and scholars.

In the field of marketing, there have been many research results on the evaluation of Customer Value. By establishing a random RFM model, Ma Baolong et al. [4] analyzed customer turnover time (R), purchase times (F) and purchase amount (M) from three dimensions to realize customer segmentation and customer value calculation. Soumaya Lamrhari[5] proposes a social CRM analytics framework that includes a variety of analytics methods aimed at improving customer retention, acquisition, and conversion. Liu et al. [6] improved the redundancy, frequency, and Maintenance Value (RFM) model for evaluating medical records using mixed packet Genetic algorithm (GGA) and density-based noise Application Spatial clustering (DBSCAN) algorithm. Cheng Dong et al. [7] believe that CLV model is the basis for enterprises to conduct customer relationship management, predicting the lifetime value of customers by predicting their consumption frequency and consumption amount.

Clustering Algorithm is an algorithm that compares the similarity of samples by using sample features and divides samples with similar attributes into the same class or cluster [8]. Representative clustering algorithms include K-means partition clustering, DBSCAN density clustering, Chameleon hierarchical clustering, spectral clustering and Gaussian mixture clustering. In recent years, faced with the location problem of large-scale customer groups, domestic and foreign scholars put forward the solution of clustering and re-location. The K-means clustering algorithm is used to determine the optimal cluster number and cluster range, and then the location is selected in these areas [9, 10]. Based on the grid model and the density of O2O takeout orders, Luo Jian et al. [11] used the grid density clustering method to divide the takeout business circle. In the face of massive dynamic customer demand, Ge Xianlong et al. [12] used historical data to predict dynamic demand from three dimensions: demand forecasting, demand clustering and demand quota. Gao Xuedong et al. [13] took customer clustering into consideration in the field of distribution center location and designed a customer clustering method for location selection limited by road network and distribution scale. Liang Xi et al. [14] built a multi-objective optimization model of logistics network cost minimization and environmental impact minimization, and first conducted clustering operations on customer points before location selection. Liu[15] constructed a new customer value portrait framework to identify industrial customer value according to different types of behavior characteristics and emerging trends of natural gas market, and combined Gower's differentiation coefficient with PAM clustering algorithm to establish a visual customer value classification model.

ISSN: 2692-7608 DOI: 10.6981/FEM.202404_5(4).0007

Most of the above studies focus on the location of distribution centers for existing customers [16], and rarely consider the role of potential customers in the location of distribution centers, ignoring this important indicator. When the K-means algorithm is used to perform cluster analysis on the longitude and latitude of customer points, potential customers and real customers are often treated equally [17]. However, in real life, potential customers and real customers have different contributions to clustering, and the possibility of each potential customer transforming into a real customer in the future is also different. In response to this problem, Fukunaga Hara et al. [18] proposed a K-means clustering algorithm based on information entropy with accurate attribute weighting to achieve higher precision and more stable clustering. Wang Zezhou et al. [19] proposed a weighting method based on bias entropy, which relaxed the normalization constraint conditions during clustering. In view of the limitations of the traditional clustering criteria based on linear coupling of information entropy and consistency coefficient, expert weights were determined according to the contribution of experts to their own categories, overcoming the shortcomings of the traditional method.

To sum up, aiming at the location problem of distribution center under the uncertain value of potential medical customers, this paper proposes a K-means clustering location method based on entropy weight method. This method is used to analyze the customer data of pharmaceutical logistics company, and the potential value of pharmaceutical customers is introduced into the cluster partition as an important index, and then the scientific and reasonable location strategy of distribution center is proposed. First of all, it is necessary to pre-process the potential customer data of pharmaceutical logistics enterprises, and then build a set of pharmaceutical customer potential value evaluation index system from five aspects: customer current value, customer development potential, customer operating cost, regional customer demand and industry competition. Then, the entropy weight method is used to calculate the weight of each index and the final comprehensive score, and the score is normalized. After screening out the customer points that the central warehouse can distribute to and remote customer points, the CKM clustering algorithm with capacity constraints is used to cluster and locate the existing customers and potential customers. The CKM algorithm was compared with other clustering algorithms by ARI and AMI internal evaluation indexes and DBI and SC external evaluation indexes to verify the clustering effectiveness .

2. Problem Description and Index System Construction

2.1. Problem Description

P Pharmaceutical Logistics Company is located in the main urban area of Chongqing, has been deeply engaged in the field of pharmaceutical supply chain services for more than ten years, is a comprehensive pharmaceutical logistics company, the main business covers general drug wholesale, new drug agent, medical equipment supply, health products and family planning products of a full range of services. As the company has only one central warehouse, which is difficult to meet the distribution needs of existing customers and potential customers, it is proposed to establish a distribution center to provide customers with drug distribution services. The distribution process diagram is shown in Figure 1.

On the basis of considering the needs of existing customers and potential customers, m distribution centers are established to provide drug distribution services for n customers. The location of the central warehouse of the enterprise has been determined in the early stage, and with the continuous expansion of the scale of enterprise customers, it is necessary to set up a distribution center according to the customer distribution and distribution volume. In the process of selecting m distribution centers, we should not only consider the needs of existing customers, but also consider the distribution needs of future potential customers with the development of enterprises. In addition, the value of existing customers and potential

ISSN: 2692-7608 DOI: 10.6981/FEM.202404_5(4).0007

customers will also change dynamically, because the construction of distribution centers is often long-term, in the location process, it is necessary to consider the value of existing customers and potential customers as much as possible to meet the needs of enterprise development. Therefore, how to scientifically locate the distribution center on the basis of fully considering the potential value of customers is particularly important.

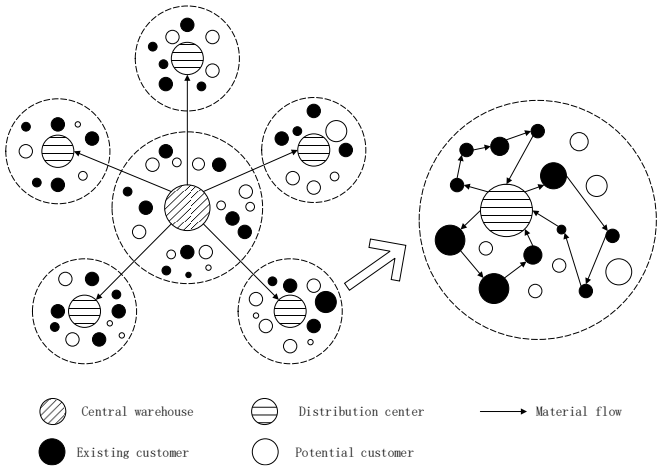

Figure 1. Distribution diagram of P pharmaceutical logistics enterprises

2.2. Customer Potential Value Evaluation Index System Construction

Whether pharmaceutical logistics enterprises can effectively evaluate the potential value of customers, the key lies in how to establish a set of scientific and reasonable evaluation index system with sufficient data support. In the process of market competition, the reputation and social influence of pharmaceutical logistics enterprises are constantly changing, and the customers they have will also change dynamically, the existing customers may be lost, and the customers of other enterprises may also become their future customers. There is a many-to-many relationship between potential pharmaceutical customers and pharmaceutical logistics enterprises, and the value of potential customers to each pharmaceutical logistics enterprise is not the same in the actual market. Therefore, in order to comprehensively evaluate the potential value of medical customers, many influencing factors should be considered comprehensively.

The potential value of pharmaceutical customers can effectively reflect the retention of existing customers and the conversion of potential customers. Different from RFM model and CLV model, which focus on the evaluation of existing customers, this paper analyzes the potential value of existing customers and potential customers in order to evaluate the potential value of pharmaceutical customers more accurately by combining the characteristics of pharmaceutical logistics distribution, such as strong timeliness, small batch, high value, wide coverage of pharmaceutical customers and strong dispersion.

According to the research of many scholars and the practice of enterprise development, the factors that affect the potential value of customers can be divided into five aspects: customer current value, customer development potential, customer operating cost, regional customer demand and industry competition. On the one hand, the contribution of pharmaceutical customers to the enterprise and their own development potential is an important aspect to reflect its potential value; On the other hand, large-scale pharmaceutical customers have higher customer acquisition costs, and enterprises will invest more in and strive for such customers, and customer acquisition costs will also affect the potential value of customers to a certain extent. In addition, under the influence of market competition, the traditional pharmaceutical logistics distribution system is difficult to meet the needs of pharmaceutical customers. In each region, pharmaceutical customers have a large demand for pharmaceutical logistics enterprises,

ISSN: 2692-7608 DOI: 10.6981/FEM.202404_5(4).0007

and the greater the potential to be transformed into real customers, and the greater the potential value.

Based on the relevant literature [4,7] and the research practice of pharmaceutical logistics enterprises, a potential value evaluation index system for pharmaceutical customers is constructed, which includes 5 secondary indexes and 20 tertiary indexes, as shown in Table 1. In the process of construction, the accessibility of data is fully considered, and the evaluation is carried out by the combination of directional index and quantitative index.

Table 1. Potential value evaluation index system table of pharmaceutical customers

Primary index	Secondary index	Three-level index
Pharmaceutical customer potential value	Customer current value B_1	C_1~C_5
	Customer development potential B_2	C_6~C_{11}
	Customer operating cost B_3	C_{12}~C_{14}
	Regional customer demand B_4	C_{15}~C_{17}
	Industry competition B_5	C_{18}~C_{20}

Table 2 describes C1 to C20 indicators and their meanings.

Table 2. Detailed list of specific indicators of potential value of pharmaceutical customers

Three-level index	Index type	Three-level index	Index type
Monthly average purchase quantity C_1	quantify	Trade intention C_{11}	qualitative
Monthly average number of purchases C_2	quantify	Direct cost of acquisition C_{12}	quantify
The latest purchase quantity C_3	quantify	Marketing input C_{13}	quantify
Last purchase time C_4	quantify	Average service price C_{14}	quantify
Customer recommended energy level C_5	qualitative	Regional population C_{15}	quantify
Customer type C_6	qualitative	Area C_{16}	quantify
Procurement growth rate C_7	quantify	Regional GDP C_{17}	quantify
Procurement balance C_8	quantify	Number of pharmaceutical logistics enterprises C_{18}	quantify
On-time payment status C_9	qualitative	Market share C_{19}	quantify
Debt service performance C_{10}	qualitative	Customer competitiveness C_{20}	qualitative

3. Method Model

3.1. Relevant Definition Assumption

Let the clustering data set be $X=\{x_i \mid x_i \in R^m, i=1,2,\cdots,n\}$, n customer groups use Ui (i=1, 2,... n) means. The centroid of n cluster customer groups is $c(U_1), c(U_2), \cdots, c(U_n)$. It is defined as follows.

ISSN: 2692-7608 DOI: 10.6981/FEM.202404_5(4).0007

Definition 1 Let's set two vectors $x_i = (x_{i1}, x_{i2}, \cdots, x_{in})$ and $y_j = (y_{j1}, y_{j2}, \cdots, y_{jn})$ representing two objects, the Euclidean distance between them is defined as follows:

$$d(x_i, x_j) = \sqrt{\sum_{d=1}^{m} (x_{id} - x_{jd})^2} \tag{1}$$

Definition 2 The centroid of a data object of the same class is defined as follows:

$$c(U_i) = \frac{1}{|U_i|} \sum_{a_j \in U_i} x_j \tag{2}$$

Where, c(Ui) represents the cluster centroid; $|U_i|$ Represents the number of data objects in the Ui.

Definition 3 The contour coefficient s for a sample is:

$$s = \frac{b - a}{Max(a, b)} \tag{3}$$

Where, a represents the average distance between a sample and other samples in the cluster, and b represents the average distance between a sample and other samples in the cluster.

The contour coefficients SK of n1 data objects are defined as follows:

$$S_k = \frac{\sum_{i=1}^{n1} s_i}{n_1} \tag{4}$$

According to definition 1, the weighted Euclidean distance dr(xi, xj) is:

$$d_r(x_i, x_j) = \sqrt{r_1 \left| x_{i1} - x_{j1} \right|^2 + \cdots + r_n \left| x_{ik} - x_{jk} \right|^2} \tag{5}$$

3.2. The Entropy Weight Method is Used to Calculate Potential Customer Value

Before applying the entropy weight method, it is necessary to define the index attributes, that is, positive and negative indicators. An increase in the positive indicator value is beneficial for decision-making objectives, such as the increase in corporate profitability, which is easier to achieve. The reduction of negative index value is also conducive to decision-making, such as cost reduction and easier realization of enterprise objectives.

The process of calculating customer potential value with entropy weight method is as follows:

(1) The three level index values are normalized:

$$Y_{ij} = \frac{x_{ij} - \min\left\{x_{1j}, \cdots, x_{nj}\right\}}{\max\left\{x_{1j}, \cdots, x_{nj}\right\} - \min\left\{x_{1j}, \cdots, x_{nj}\right\}} \tag{6}$$

ISSN: 2692-7608 DOI: 10.6981/FEM.202404_5(4).0007

The negative indicator is calculated as follows:

$$Y_{ij} = \frac{\max\left\{x_{1j}, \cdots, x_{nj}\right\} - x_{ij}}{\max\left\{x_{1j}, \cdots, x_{nj}\right\} - \min\left\{x_{1j}, \cdots, x_{nj}\right\}} \tag{7}$$

Where xij represents the original value of the i customer under the j index, and n represents the number of all potential customers;

(2) Calculate the proportion of the index value Pij, which represents the proportion of the j indicator of the i customer in the sample value of the indicator, and use it as the probability when calculating information entropy.

$$P_{ij} = \frac{Y_{ij}}{\sum_{i=1}^{n} Y_{ij}} (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots,\ f) \tag{8}$$

Where, f represents the number of indicators; Yij represents the non-dimensional standardized value of the i customer under the j indicator.

(3) The information entropy matrix of each potential pharmaceutical retailer customer is calculated. The formula for calculating the index entropy is as follows:

$$E_j = -\ln(n)^{-1} \sum_{i=1}^{n} p_{ij} \ln p_{ij} (0 \le E_j \le 1) \tag{9}$$

(4) The weight of each index affecting the potential value of medical customers is calculated by calculating the information redundancy degree.

$$d_j = 1 - E_j, (j = 1, 2, \ldots, m) \tag{10}$$

$$W_j = \frac{d_j}{\sum_{j=1}^{m} d_j}, (j = 1, 2, \ldots, m) \tag{11}$$

Where 1−Ej is the entropy redundancy.

(5) Calculate the combined score of each customer sample, that is, customer value.

$$S_i = \sum_{j=1}^{m} w_j P_{ij}, (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m) \tag{12}$$

(6) The comprehensive score was normalized.

The specific function of normalization is to obtain the value between [0, 1]. The larger the original comprehensive score is, the larger the result after normalization will be. In other words, the higher the comprehensive score of a medical customer is, the greater the potential value of the customer.

ISSN: 2692-7608 DOI: 10.6981/FEM.202404_5(4).0007

3.3. CKM Clustering Algorithm Determines the Cluster Customer Base

K-means clustering algorithms usually use Euclidean distance to measure the similarity between data objects. The smaller the Euclidean distance between two objects, the greater their similarity; On the contrary, the similarity is smaller. The CKM clustering algorithm proposed in this paper is based on the original K-means clustering algorithm and gives weight to Euclidean distance considering the potential value of medical customers. Assuming that the weight of the i potential customer is ri, the attribute of customer potential value is added to the original attribute of customer latitude and longitude, and the clustering is appropriately scaled up and down according to the weight of ri. This makes customers with high potential value play a greater role in clustering, while customers with low potential value play a smaller role in clustering, which truly reflects the role of each customer value in actual clustering.

The K-means clustering algorithm is a statistical analysis method that boils the homogeneous continuous variables together to form different clusters through an iterative process. Because of its simple principle and high efficiency, the algorithm is widely used, but the clustering result is easily affected by other factors, such as the determination of k value, the selection of outliers and the initial clustering center. In order to reduce the influence of outliers on the clustering effect, elbow method or contour coefficient is usually used to determine the best k value and remove the isolated points before clustering. The CKM clustering algorithm adds lead value as an attribute to the capacity-constrained K-means clustering. Bring customer potential value ri into equation (5) for calculation. Using Python to write CKM clustering algorithm, considering the needs of potential customers, match the dual attributes of latitude and longitude and potential value for each medical customer, and set the maximum demand constraint for each cluster, so as to limit the size of the cluster customer base.

The CKM clustering algorithm needs to input the longitude, latitude and potential value of cluster objects as data sets, and determine the maximum capacity constraint d-limit according to enterprise requirements. Then, k objects are randomly selected as the initial clustering center, and the minimum k traversal is performed, where $K = sum(demand)/d\_limit$. In this process, according to the traditional K-means clustering algorithm, each cluster object is assigned to the most similar (closest) cluster, and the minimum mean point in each cluster is recalculated as the new cluster center. On this basis, a new step is added to calculate the contour coefficient under each k value to evaluate the clustering effect. The value range of contour coefficient is [-1, 1]. The closer the distance between samples of the same category and samples of different categories, the higher the score. If the clustering effect is not good, a larger k value is selected until the k value with the best clustering effect is selected. After the centers of K clusters do not change, the optimal number of clusters and clustering results are output.

4. Case Analysis

In order to verify the effectiveness of the algorithm and model, this paper selects the data of existing customers and potential customers provided by P pharmaceutical logistics company for research, calculates the potential value customers of customers, and considers them into the location of distribution center. The experiment was carried out in two steps. The first step was to verify the effectiveness of the proposed CKM clustering algorithm through four clustering effectiveness evaluation indicators; the second step was to apply the algorithm to the complete data set to find out the final cluster customer base and the location of the distribution center.

ISSN: 2692-7608 DOI: 10.6981/FEM.202404_5(4).0007

4.1. Data Processing

4.1.1. Longitude and Latitude Data Processing

Since the customer information provided by the enterprise is the geographical location of each medical customer, it is also necessary to convert all customer addresses into latitude and longitude to provide data support for the subsequent algorithm. The steps for converting an address to latitude and longitude are as follows:

(1) Save the geographical location data of potential customers in excel tables, clean incomplete data, and delete duplicate data rows;

(2) Apply for a key in the Amap console;

(3) Call the Amap API to write python code, convert the geographical location of potential customers into latitude and longitude, so as to determine the horizontal and vertical coordinates of the demand point.

4.1.2. Evaluation Index Calculation

The three indexes in the above evaluation index system are classified, and there are 15 positive indicators: Monthly average purchase volume, monthly average purchase frequency, last purchase volume, customer recommendation level, customer type, purchase growth rate, procurement balance, on-time payment, debt service performance, willingness to re-trade, regional population, regional area, regional GDP, number of pharmaceutical logistics enterprises, customer competitiveness; There are five negative indicators: time since the last purchase, direct cost of customer acquisition, marketing investment, average service price, and market share.

The entropy weight method was used to calculate the weight of each indicator of medical customers, as shown in Table 3.

Table 3. Weight of indicators of pharmaceutical customers

Three-level index	Index type	Three-level index	Index type
Monthly average purchase quantity C_1	0.137	Trade intention C_{11}	0.028
Monthly average number of purchases C_2	0.155	Direct cost of acquisition C_{12}	0.015
The latest purchase quantity C_3	0.142	Marketing input C_{13}	0.017
Last purchase time C_4	0.135	Average service price C_{14}	0.004
Customer recommended energy level C_5	0.028	Regional population C_{15}	0.018
Customer type C_6	0.027	Area C_{16}	0.02
Procurement growth rate C_7	0.014	Regional GDP C_{17}	0.024
Procurement balance C_8	0.142	Number of pharmaceutical logistics enterprises C_{18}	0.004
On-time payment status C_9	0.028	Market share C_{19}	0.012
Debt service performance C_{10}	0.028	Customer competitiveness C_{20}	0.021

The results in Table 3 show that the weight value of the average monthly purchase times C2 is the largest, and the weight of the average service price C14 and the weight of the pharmaceutical wholesale enterprise C18 is the least. It can be seen that the current value of the customer is the most important factor affecting the potential value of the customer.

ISSN: 2692-7608 DOI: 10.6981/FEM.202404_5(4).0007

The comprehensive score of each medical customer is calculated based on the weight of each indicator. The calculation formula is as follows:

$$V_i = C_{1i} \times W_1 + C_{2i} \times W_2 + \cdots + C_{20i} \times W_{20} \tag{13}$$

In the formula, Vi represents the comprehensive score of medical customer i; $C_{1i}, C_{2i}, \cdots C_{20i}$ represent the normalized values of 20 three-level indicators; $W_1, W_2, \cdots, W_{20}$ represent the weights of 20 three-level indicators.

The comprehensive score was normalized and the potential value ri of each medical customer was obtained.

4.1.3. Clustering Data Set Determination

Before K-means clustering, various complex noises and outliers in the data set need to be processed to improve the effectiveness and robustness of the clustering algorithm. The packaged DBSCAN clustering algorithm in Matlab is called, the pre-processed data is inserted, the two parameters Eps and MinPts are adjusted according to the requirements, and the customer base and outlier points of the distribution center need to be established are output. In this example, after several experimental tests and data verification, the coverage radius of the distribution center of the pharmaceutical logistics enterprise is set to 50km, and the neighborhood radius Eps is set to half of the coverage radius, that is, 25km. According to the geographical coordinate system, the actual distance corresponding to the longitude difference of 1 degree in this region is about 89km, and the actual distance corresponding to the latitude difference of 1 degree is about 111km. Based on these parameters, the approximate value of neighborhood radius Eps on coordinates is further calculated as 0.2. In addition, to ensure the validity and accuracy of clustering, the minimum density value MinPts is set to 25.

The original central warehouse of the pharmaceutical logistics company radiates to a certain range, and the customer points within 30 kilometers of the direct distance from the central warehouse are directly distributed by the warehouse with reference to the opinions of the management personnel with rich practical experience. After sifting out the customers delivered by the central warehouse, a cluster dataset of 38,906 existing and potential customers was finalized.

4.2. Data Processing

In order to verify the effectiveness of the CKM clustering algorithm, 1000 relevant data in Section 4.1 were selected as the training set, and the results of the two-dimensional K-means clustering algorithm (only considering the latitude and longitude of the customer) were selected as the standard clustering label. Matlab R2020a and Spyder (Python3.9) were selected as the programming tools in the experiment. By comparing the CKM clustering algorithm with the three-dimensional K-means clustering algorithm (considering customer latitude and longitude and customer value), the system clustering algorithm and the genetic algorithm, the clustering algorithm proposed in this paper can be judged to be superior to the comparison algorithm by comparing and observing the size of the clustering effectiveness evaluation index when the four algorithms are in $2 \le k \le 10$.

There are many kinds of clustering quality evaluation indexes. This paper selects 4 kinds of commonly used clustering effectiveness indexes for simulation. The two external evaluation indicators are Adjusted Rand Index (ARI) and Adjusted Mutual Information Index (AMI), and the Silhouette Coefficient (Silhouette Coefficient). SC, Davidson-Bouldin Index (DBI) two internal evaluation indicators. Both ARI and AMI are used to measure the similarity between the clustering result and the real category, and the value range is [-1, 1]. However, ARI only considers the influence of random allocation, while AMI also considers the influence of random

ISSN: 2692-7608 DOI: 10.6981/FEM.202404_5(4).0007

allocation and category imbalance. In both cases, the larger the value, the more consistent the clustering result is with the real situation. SC is an evaluation method for the quality of clustering effect. Combining the two factors of cohesion and separation, SC can be used to evaluate the influence of different algorithms or different operation modes of algorithms on clustering results on the basis of the same original data. Its value is between [-1, 1], and the closer to 1 indicates that both cohesion and separation are relatively better. DBI, also known as classification accuracy index, calculates the sum of the average distance between any two categories of intra-class distance divided by the distance between the center of the two clusters to find the maximum value. The smaller the DBI value, the better the clustering effect.

Figure 2-5 shows the statistical results of the evaluation indicators of the four algorithms in the data set. For external indicators ARI and AMI, the four algorithms show similar trends in the data set, but CKM clustering algorithm is superior to the other three clustering algorithms on the whole. When k=2, CKM clustering algorithm achieves the most obvious optimization effect, and its ARI value is increased by 61.9% compared with systematic clustering method. Compared with the three-dimensional K-means clustering algorithm, the improvement is 11.4%, and the improvement is 16.1% compared with the genetic algorithm. The AMI value increased by 51.9%, 11% and 15%, respectively. For the other two external indicators DBI and SC, the clustering effect of CKM clustering algorithm is stable and the performance is optimal, and the effect of the comparison algorithm is significantly different when taking different k values. With the increase of K value, SC value decreases and DBI value increases. Because the smaller DBI index is, the better it is, it can be seen that the clustering effect of CKM clustering algorithm deteriorates with the increase of k value. In summary, the clustering validity index of the CKM clustering algorithm proposed in this paper is generally better than that of the comparison algorithm on this dataset, which indicates that the CKM clustering algorithm is more applicable on this dataset. In general, the data can be clustered into 2 categories with the largest ARI, AMI and SC and the smallest DBI.

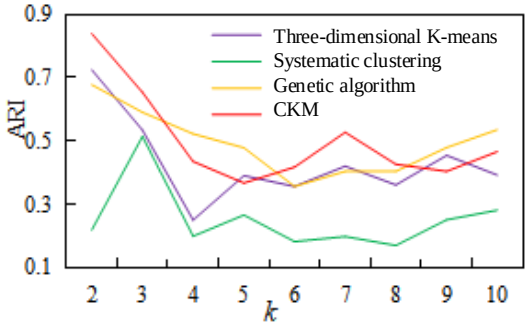

Figure 2. Evaluation results of the Adjusted Rand Index (ARI)

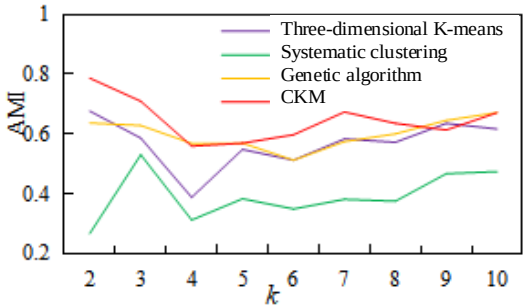

Figure 3. Evaluation results of the Adjusted Mutual Information Index (AMI)

ISSN: 2692-7608 DOI: 10.6981/FEM.202404_5(4).0007

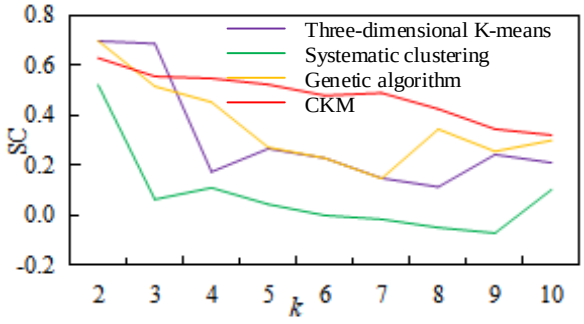

Figure 4. Evaluation results of the Silhouette Coefficient (SC)

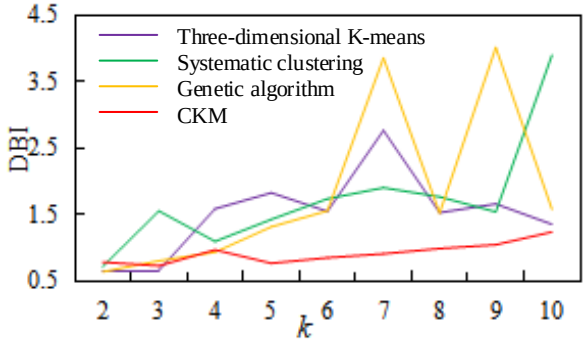

Figure 5. Evaluation results of the Davies-Bouldin Index (DBI)

4.3. Algorithm Application

After verifying the effectiveness of CKM algorithm, the location of distribution center is determined by taking P pharmaceutical logistics enterprise as an example. Cluster the cluster data set obtained in the previous section. The data set fields include "Longitude of medical customer", "Latitude of medical customer" and "potential value of medical customer", in which the potential value of medical customer is ri and the maximum capacity limit of distribution center is 800. The contour coefficient obtained by running Python code is shown in Figure 6. When k=6, the total contour coefficient value is the highest, and the clustering effect is the best at this time.

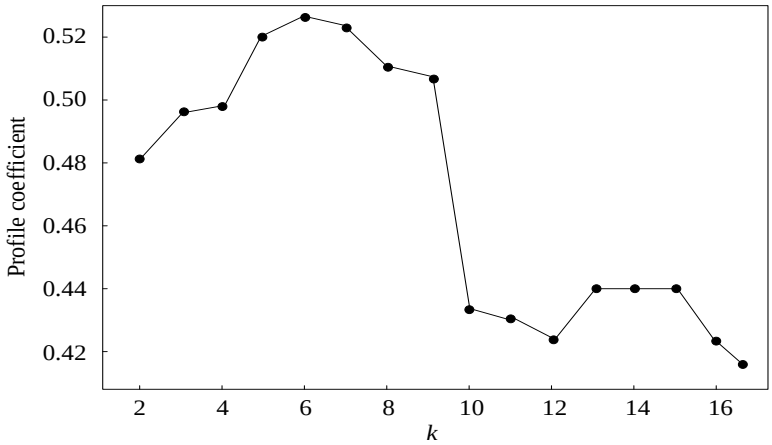

Figure 6. Cluster silhouette coefficient

The CKM clustering results obtained by running Python are shown in Figure 7.

ISSN: 2692-7608 DOI: 10.6981/FEM.202404_5(4).0007

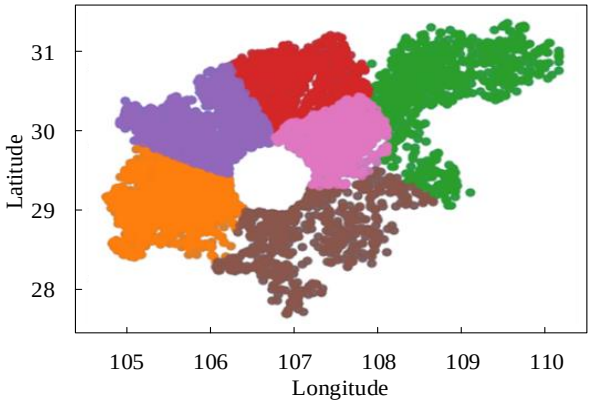

Figure 7. Clustering result graph

The clustering center obtained by CKM clustering is the location of the distribution center considering the potential value of medical customers. The location of the distribution center of each cluster is shown in Table 4.

Table 4. Distribution center site selection status

Cluster customer base	Number of customers	Distribution center location
1	6524	107.338768, 30.049761
2	8128	105.519722, 29.201691
3	4575	106.970596, 28.715919
4	5658	106.964729, 30.777491
5	7385	105.934819, 30.158813
6	6636	108.560432, 30.78524

5. Conclusion

In order to solve the location problem of distribution center under the condition of uncertain customer value, a K-means clustering location method based on entropy weight method was proposed, and the potential value of medical customer was introduced into the location scheme as an important attribute. CKM clustering algorithm is used to realize clustering considering the potential value of customers and make up for the shortcomings of predecessors in the field of medical logistics location. This research has important reference significance for the establishment of a unified, orderly and reasonable distribution center for medical logistics. Compared with the traditional method of customer value calculation, the method proposed in this paper is more suitable for large-scale customer value calculation, and makes full use of the clustering difference degree of each attribute of the data object, so that the clustering result is closer to the actual classification of the data object. In the future research, it will be carried out from the following aspects: First, the entropy weight method and K-means algorithm are further combined to reduce the work intensity and meet the needs of large-scale customer data value evaluation and clustering partition; Secondly, the paper explores the influence of potential customer value on the location selection of regional distribution center, in order to put forward a new location selection model.

Acknowledgments

This work is supported by the Chongqing Technology Innovation and Application Development Project (CSTB2022TIAD-GPX0015) and the Chongqing Social Science Planning Project (2021ZDZK14).

ISSN: 2692-7608 DOI: 10.6981/FEM.202404_5(4).0007

References

[1] MENG H M. Development Strategies for Third-Party Pharmaceutical Logistics Industry in the Context of New Version GSP [J]. China Logistics and Purchasing, 2018, 561(20): 76.

[2] YANG Y L, ZHANG J, SUN W J, et al. Location-Path Problem Based on Pharmaceutical Prepositioned Warehouses Under Dynamic Demand [J]. Control and Decision, 2023, 38(6): 1670-1678.

[3] Gopalakrishna Srinath, Crecelius Andrew T, Patil Ashutosh. Hunting for new customers: Assessing the drivers of effective salesperson prospecting and conversion [J]. Journal of Business Research, 2022, 149: 916-926.

[4] MA B L, LI F, WANG G, et al. The Random RFM Model and Its Application in Retail Customer Value Recognition [J]. Journal of Industrial Engineering and Engineering Management, 2011, 25(1): 102-108.

[5] Lamrhari Soumaya, Ghazi Hamid El, Oubrich Mourad, et al. A social CRM analytic framework for improving customer retention, acquisition, and conversion [J]. Technological Forecasting and Social Change, 2022, 174: 121275.

[6] Liu Yishu, Chen Chen. Improved RFM Model for Customer Segmentation Using Hybrid Meta-heuristic Algorithm in Medical IoT Applications[J]. International Journal on Artificial Intelligence Tools, 2022, 31(1).

[7] CHENG D, SUN Y L, XUE W. Robust Measurement of Non-Contractual Customer Lifetime Value: A Comprehensive Study of Classical Methods and Machine Learning Algorithms [J]. Management Review, 2019, 31(4): 83-98.

[8] WANG M, SONG X H, LIU Y, et al. Neural Tangent Kernel K-Means Clustering [J]. Journal of Computer Applications, 2022, 42(11): 3330-3336.

[9] LI T Y, LV X N, LI F, et al. Optimization Model and Algorithm for Food Delivery Routing Considering Dynamic Demand [J]. Control and Decision, 2019, 34(2): 406-413.

[10] WANG Y, HUANG S Q, LIU Y, et al. Optimization Study of Logistics Multi-Distribution Center Location Selection Based on K-means Clustering Method [J]. Journal of Highway and Transportation Research and Development, 2020, 37(1):141-148.

[11] LUO J, TANG J F, YU Q Y, et al. Boundary Delineation of O2O Food Delivery Service and Discovery of Customer Demand Distribution Patterns [J]. Chinese Journal of Management Science, 2023, 31(3): 58-68.

[12] GE X L, WEN P Z, XUE G Q. Two-Level Dynamic Delivery Routing Optimization Based on Demand Prediction [J]. Chinese Journal of Management Science, 2022, 30(8): 210-220.

[13] GAO X D, GU S J, BAI C, et al. Customer Clustering Algorithm Considering Logistics Delivery Network Structure and Delivery Volume Constraints [J]. Systems Engineering-Theory and Practice, 2012, 32(1): 173-181.

[14] LIANG X, KAI W. Two-Level Closed-Loop Logistics Network Location-Routing Optimization considering Customer Clustering and Product Recovery [J]. Journal of Computer Applications, 2019, 39(2): 604-610.

[15] Liu S, Gong C, Pan K. A combinatorial model for natural gas industrial customer value portrait based on value assessment and clustering algorithm[J]. Frontiers in Energy Research, 2023, 11: 1077266.

[16] NI W H, CHEN T. Emergency Logistics Distribution Center Location Selection Based on Clustering - Centroid Method [J]. Journal of Nanjing Tech University(Natural Science Edition), 2021, 43(2): 255-263.

[17] ZHAO Z Q, ZHANG L T, WANG W C, et al. Location Selection of Fresh Agricultural Products Forward Warehouse based on Customer Demand Distribution [J]. Computer Applications and Software, 2021, 38(10): 107-113, 124.

[18] YUAN F Y, ZHANG X C, LUO S B. Exact Attribute Weighting K-means Clustering Algorithm based on Information Entropy [J]. Journal of Computer Applications, 2011, 31(6): 1675-1677.

[19] WANG Z Z, CHEN Y X, XIANG H C. A Study on an Improved Expert Fuzzy C-Means Clustering Weighting Method [J]. Chinese Journal of Management Science, 2021, 29(2): 177-183.